

# Detection and Resolution of References to Meeting Documents

Andrei Popescu-Belis<sup>1</sup> and Denis Lalanne<sup>2</sup>

<sup>1</sup> University of Geneva,  
School of Translation and Interpretation (ETI), TIM/ISSCO,  
40, bd. du Pont d'Arve,  
CH-1211 Geneva 4, Switzerland  
[andrei.popescu-belis@issco.unige.ch](mailto:andrei.popescu-belis@issco.unige.ch)

<sup>2</sup> University of Fribourg,  
Faculty of Science,  
DIUF/DIVA,  
3, ch. du Musée,  
CH-1700 Fribourg, Switzerland  
[denis.lalanne@unifr.ch](mailto:denis.lalanne@unifr.ch)

**Abstract.** This article describes a method for document/speech alignment based on explicit verbal references to documents and parts of documents, in the context of multimodal meetings. The article focuses on the two main stages of dialogue processing for alignment: the detection of the expressions referring to documents in transcribed speech, and the recognition of the documents and document elements that they refer to. The detailed evaluation of the implemented modules, first separately and then in a pipeline, shows that results are well above baseline values. The integration of this method with other techniques for document/speech alignment is finally discussed.

## 1 Introduction

Documents are often the main support for communication in group meetings. For instance, slides are used for talks, and are generally displayed in sequence, being thus naturally aligned with the presenter's utterances. This is not the case, however, when the supporting documents are not so obviously set into focus, for instance when reports or articles are discussed during a meeting.

When meetings are recorded and stored in a database that can be accessed by a meeting browser, it is necessary to detect the temporal alignment between speech and documents or sub-document elements. This kind of alignment has to be derived from the linguistic content of speech and documents, and from clues in other modalities.

We study in this paper the alignment of transcribed speech and electronic documents, based on the references that are made explicitly in speech, such as "the title of our latest report" or "the article about . . .". A number of processing modules required to carry out this task are described in Section 2, and techniques

for document structuring are briefly outlined (2.2). Section 3 defines reference-based document/speech alignment, then describes the proposed methods for the detection of expressions referring to documents and the recognition of the document elements they refer to. The press-review meetings used in this experiment and the evaluation methods that we designed are described in Section 4. Results appear in Section 5. Finally, the place of reference-based alignment among other document/speech alignment techniques is discussed in Section 6.

## 2 Document/Speech Alignment for Meeting Browsing

Meeting processing and retrieval applications target several types of users. For instance, a professional who missed a meeting could use such an application to browse through the meeting's content directly to the most relevant points, without viewing or listening to the entire recording. Likewise, someone who attended a meeting but who would like to review some points, such as the decisions that were made, could benefit from a meeting browser, as well as someone who would like to track the progress of issues over several meetings. Once an episode of interest has been spotted in a meeting, a meeting browser should allow the user to explore the transcript, or to watch/listen to the episode, or to check the documents that were discussed.

### 2.1 Importance of References to Documents for Meeting Browsing

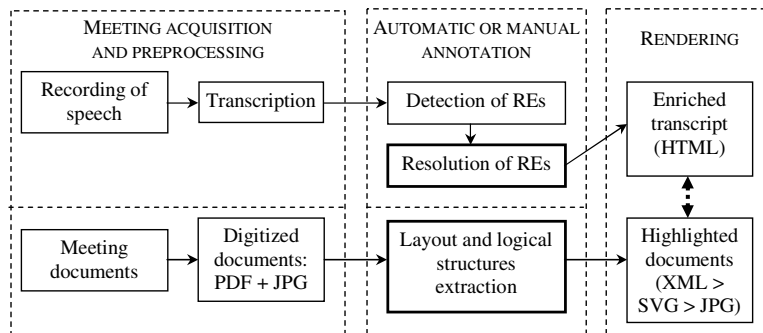
When meetings deal with one or several documents, it becomes important to align in a precise manner each episode of the meeting to the sections of the documents that are discussed in it, and vice-versa. This allows a meeting browser to retrieve the episodes of a meeting in which a particular section of a document was discussed, so that the user can find out what was said about it. Conversely, the application can also display the documents relevant to a given episode of a meeting, while the user browses through that episode. A study of user requirements has shown that queries frequently involve information related to meeting documents [1].

The references made in speech to the meeting documents are a fined-grained type of information that allows document/speech alignment. Using these references, the multimodal rendering of the meeting can be enhanced as shown in Fig. 1. The expressions that refer to documents are coded, in this implementation, as hyperlinks towards the right part of the window: clicking on such a link highlights the article referred to by that expression. This approach can of course be integrated to larger, more complex browsers.

The resolution of references to documents is a cross-channel task that enhances dialogue and document browsing. The task requires significant pre-processing of data (Fig. 2). The most significant tasks are: the generation of a transcript of the utterances produced by each speaker; the generation of an abstract representation of each document structure; the detection of the expressions from the transcripts that refer to meeting documents; and the identification of the document element each of these expressions refers to. The latter two tasks are the main object of this chapter.



**Fig. 1.** Aligned browsing of meeting transcript and documents. Clicking on a referring expression (underlined) highlights the corresponding document element.



**Fig. 2.** Components of an application for the resolution of references to documents

## 2.2 Construction of the Logical Structure of Documents

The PDF format has become very common for disseminating nearly any kind of printable documents, since it can be easily generated from almost every other document format. However, because its use is limited to displaying and printing, its value for retrieval and extraction is considerably reduced. Our experience has shown that the reading order of a text is often not preserved, especially in documents having a complex multi-column layout, such as newspapers. Even recent tools that extract the textual content of PDF documents do not reveal the physical and logical structures of documents. To overcome these limitations, we designed and implemented Xed, a tool that reverse engineers electronic documents and extracts their layout structure [2]. This approach merges low-level text extraction methods with layout analysis performed on synthetically generated TIFF images. Xed has been tested with success on various document classes with complex layouts, including newspapers.

In the present study, we consider that newspaper front pages have a hierarchical structure. The following elements (in *Typewriter* font) are used. A *Newspaper* front page bears the newspaper's Name, the Date, one *Master Article*, zero, one or more *Highlights*, one or more *Articles*, etc. Each con-

tent element has an ID attribute bearing a unique index. An **Article** is composed of a **Title**, a **Subtitle**, a **Source**, the **Content** (mandatory), and one or more **Authors** and **References**.

To obtain data with 100% correct document structure for the application to document/speech alignment, the XML document segmentations have been validated manually according to the structure mentioned above, encoded in a DTD. Information about the layout structure, i.e. the bounding boxes of each logical block, topological positions, fonts, etc., was stored in separate annotation files, using pointers to the ID attributes of the logical blocks.

### 3 Reference-Based Document/Speech Alignment

#### 3.1 What Are References to Documents?

From a cognitive point of view, speakers use *referring expressions (REs)* to specify the entities about which they talk, or more accurately the representations of entities in the speaker's mind. When speakers discuss one or more documents, as in press-review meetings, they often refer explicitly to documents or various parts of documents (e.g. 'the title', 'the next article', etc.).

Reference resolution amounts to the construction of links between each RE and the corresponding document element. For example, if a speaker says: "I do not agree with the title of our latest report", then 'our latest report' refers to a paper or electronic document, and 'the title of our latest report' refers precisely to its title, an element that can be retrieved from the document structure.

Two important notions are *coreference* and *anaphora*.  $RE_1$  and  $RE_2$  are coreferent if they refer to the same entity, here a document element.  $RE_2$  is an anaphor with respect to  $RE_1$  if the element it refers to cannot be identified without making use of  $RE_1$ , then called the antecedent of  $RE_2$ . In the following example, 'the first article' is the antecedent and the pronoun 'it' is the anaphor: "The first article is particularly relevant to our company. It discusses ...". Note that anaphora may occur without coreference, as is the case with 'the first chapter' and 'the title' in this example: "The first chapter is nicely written. The title suggests that ...". The resolution of references to documents offers the advantage of a restricted set of candidate entities, when compared to anaphora or coreference resolution [3-6].

#### 3.2 The Detection of REs

The reference resolution process has in our view two main stages: (1) the detection of the REs that refer to documents; (2) the identification of the document and document element that each RE refers to. In a preliminary study [7], only the second stage could be automated: no results were available for the entire process. We present here an automated solution for the first stage as well, and evaluate the accuracy of the two combined stages.

We designed a grammar-based component that spots the REs referring to documents in the transcript of meeting dialogues (in French). We chose to consider

a manual speech transcript because an automatic one would contain too many recognition errors, which would make the evaluation of our alignment impossible. Each channel is segmented into utterances following the SDA.XML format used in our project [8]. We used the CLaRK XML environment [9]<sup>1</sup> to write a tokenizer and a grammar.

In order to detect REs that refer to documents, we created a set of pattern matching rules applying to the words of the utterances, with sometimes a left or a right context. The challenge in writing the detection grammar was to combine a priori linguistic knowledge about the form of REs with the empirical observations on our corpus<sup>2</sup>, summarized elsewhere [7]. The resulting grammar has about 25 pattern matching rules, but since most of them contain one or more logical disjunctions and optional tokens, they are equivalent to several hundred possible REs. Another challenge was to tune the coverage of the grammar to avoid too many false positives or true negatives, corresponding respectively to precision and recall errors for the RE detection task (see 4.2).

The main improvement that should be made to this method – apart from increasing the coverage and accuracy of the grammar – is the intrinsic ambiguity of certain REs, which may or may not refer to documents, depending on their context. A typical example are pronouns such as ‘it’ and indexicals such as ‘this’ or ‘this one’, which seem to require some knowledge of their antecedent in order to be tagged as referring to documents or not. A possible solution would be to develop a classifier for this task, based on surface features present in the left and right contexts and surrounding REs, or to extend the above grammar to filter out pronouns that cannot refer to documents. In the meanwhile, we tested several pattern matching rules, and kept the ones that increased recall without reducing precision too much. The failure to detect the pronouns is, however, quite penalizing for the document/speech alignment task, shown in Section 5.3.

### 3.3 The Recognition of References to Documents

Once the REs are detected, the second task is to recognize to which document and document element each RE refers, among the set of potential referents that is derived from the document structure. A first idea is to consider co-occurrences of words between the RE and the documents. For each RE, its words and the words surrounding it are matched using the cosine metric with the bag of words of each document element: **Title**, **Author**, **Content**, etc. The most similar document element could be considered as the referent of the RE, provided the similarity value exceeds a fixed threshold.

The theories of reference resolution emphasize, however, the importance of keeping track of the referents that were mentioned, in particular of the “current” referent [10]. We integrated therefore this important feature and the word-based

<sup>1</sup> Available at: <http://www.bultreebank.org/clark/>.

<sup>2</sup> For instance, most of the references are made to entire articles, using REs such as ‘the article’, ‘the [first/last] article’, ‘a short article about . . .’, or ‘the front page of Le Monde’. These examples are translated from French; ‘Le Monde’ is the name of a French newspaper.

comparison into a more complex algorithm which processes anaphoric and non-anaphoric REs differently.

The resulting algorithm processes the REs in sequence. First, it determines the document referred to by each RE, among the list of documents associated to the meeting. The criterion is that REs that make use of a newspaper's name are considered to refer to the respective newspaper, while all the other ones are supposed to refer to the current newspaper<sup>3</sup>.

The algorithm then attempts to determine the document *element* that the current RE refers to. It first decides whether the RE is anaphoric or not by matching it against a list of typical anaphors for document elements (e.g. 'the article' or 'it'). If the RE is anaphoric (and not the first RE of the meeting), then its referent is the current document element. If the RE is not anaphoric, then co-occurrences of words are used as above to find the document element it refers to: the words of the RE and the surrounding ones are matched with document elements; the one that scores the most matches is considered to be the referent of the RE. Then, the 'current document' and the 'current document element' (a single-level focus stack [10]) are updated, before processing the next RE.

Several parameters govern the algorithm, in particular the relative importance of the various matches between words of the RE and of its left/right context, with the words from document elements. Another parameter is the span of the left and right contexts, that is, the number of preceding and following words and utterances considered for matching. These parameters are tuned empirically in Section 5.2.

## 4 Data and Evaluation

The data was recorded in the document-centric meeting room set up at the University of Fribourg. Several modalities related to documents were recorded, thanks to a dozen cameras and eight microphones. These devices are controlled and synchronized by a meeting capture and archiving application, which also helps the users to organize the numerous data files [11].

In this study, we use 22 press-review meetings of ca. 15 minutes each, recorded between March and November 2003, in which participants discuss the front pages of one or more newspapers of the day, in French<sup>4</sup>. Each participant introduces one or more articles. For each article, a short monologue is followed by a brief discussion. The meetings were manually transcribed using the Transcriber tool<sup>5</sup> and exported as XML files. The structure of the 30 documents (front pages, cf. Section 2.2) was also encoded into XML files.

### 4.1 Annotation of Ground Truth REs and References

The annotation model for the references to documents was described in an earlier paper [7]. The main idea is to separate the annotation of REs from the annotation of the references to documents. REs are tagged on the XML transcript

<sup>3</sup> This method does not handle complex references such as 'the other newspaper'.

<sup>4</sup> Available at: <http://diuf.unifr.ch/im2/data.html>.

<sup>5</sup> Available at: <http://www.etca.fr/CTA/gip/Projets/Transcriber>.

using an opening `<re ID="...">` and a closing `</re>` tag. The documents and elements they refer to are encoded in a separate block at the end of the XML transcript, as links between the index of the RE (ID attribute), a document filename, and an XPath designation of the document element referred to, in the XML representation of the document structure.

In a first pass, the annotators marked the REs using their own understanding of references to documents. The most litigious cases were the impersonal references to the creator of an article, such as (in English) “*they say that ...*”. We assumed this was a reference to the author of the article, or at least to the entire article (the actual scoring procedure allows this flexibility). REs that correspond only to quotations of an article’s sentences were not annotated, since they refer to entities mentioned by the documents, rather than to the document elements.

A total of 437 REs were annotated in the 22 meetings of the corpus. This number is not due to the subjects being instructed to refer more often to documents, but is due to the document-centric meeting genre.

In a second pass, the annotators were instructed to code, for each RE, the name of the document and the XPath to the respective document element, using the templates that were generated automatically after the first pass. Examples of XPath expressions were provided. When in doubt, annotators were instructed to link the RE to the most general element, that is, the article or front page.

Inter-annotator agreement for the second pass [7], with three annotators on 25% of the data, is 96% for document assignment and 90% for document element assignment (see evaluation metric below). After discussion among annotators, we reached 100% agreement on documents, and 97% agreement on elements.

## 4.2 Evaluation of RE Detection

The evaluation of the first processing stage, RE detection, is done by comparing the correct REs with those found automatically, using precision and recall. To apply these metrics, two problems must be solved. First, to what extent is some variability on the RE boundaries tolerated? And second, how are embedded REs processed?

We consider here that the detection of only a fragment of an RE counts the same as the detection of the entire RE, i.e. a correct hit is counted if the `<re>` and `</re>` tags found by the RE detector are identical to, or comprised within the correct ones. This is somewhat similar to the MUC-7 guidelines [4], with the difference that here, no minimal fragment is required for an RE. This indulgent scoring procedure is due to the nature of our application: detecting only a fragment of an RE is indeed sufficient for document/speech alignment, if the fragment is correctly linked to a document.

Embedded REs correspond in general to embedded NPs, such as “[the title of [the next article]]” (non-embedded but intersecting REs seem to be ruled out by the recursive nature of syntax). The difficulty in scoring embedded REs is related to the above decision to score RE fragments. If only exact matches counted as correct, there would be no risk of confusion between embedded REs. But because RE fragments count as well, one should avoid counting them more than once. For in-

stance, if the RE detector generates the following markup: “the title of the first `</re>chapter</re>`”, then “chapter” should count either as a match for “the first chapter” or for “the title of the first chapter”, but not for both REs.

We propose therefore the following error counting algorithm, which loops through all the correct REs in sequence (for embedded REs, it starts with the deepest one). For each correct RE, if the system has tagged it, or has tagged an RE included it, then no error is counted, and this RE is removed from the set of system REs; if it hasn’t, count one recall error. When all correct REs have been thus tested, *recall error* is the number of recall errors that were counted, divided by the total number of correct REs. *Precision error* is the number of system REs remaining in the list (that is, not matching correct ones), divided by the total number of REs tagged by the system.

### 4.3 Evaluation of RE Resolution

If the resolution of REs is attempted on the correct set of REs, then its evaluation is done simply in terms of correctness or accuracy [7]. For each RE the referent found by the system is compared with the correct one using three criteria, and then three global scores are computed. The first one is the number of times the document is correctly identified. The second one is the number of times the document element at the `Article` level (characterized by its ID attribute) is correctly identified. The third one is the number of times the exact document element (characterized by its full XPath) is correctly identified. These values are normalized by the total number of REs to obtain scores between 0 and 1. The third metric is the most demanding one. However, we will use only the first two, since our resolution algorithms do not target sub-article elements yet.

When the resolution of REs is combined with their recognition, the evaluation method must be altered so that it does not count wrongly-detected REs, which are necessarily linked to erroneous document elements, since these are evaluated by the precision score at the level of RE detection. The method must however count the REs that were not detected (to count the missing links) and examine the detected RE fragments, which may or may not be correctly linked to documents.

We used the algorithm that scores RE detection (Section 4.2) to synchronize the indexes of the detected REs with the correct ones. This allows us to compute the three accuracy scores as defined above. These adapted metrics of the accuracy of RE resolution thus take partially into account the imperfect RE detection, but they are not influenced by detection “noise”. Therefore, to evaluate the combined process of detection and resolution, the scores for RE detection are still required.

## 5 Results

### 5.1 Scores for the Detection of REs

The grammar for the detection of REs is evaluated in terms of recall ( $R$ ), precision ( $P$ ) and f-measure ( $f$ ). The initial grammar based on prior knowledge and on corpus observation reaches  $R = 0.65$ ,  $P = 0.85$  and  $f = 0.74$ .



Experimental analysis can help to assess the value of certain rules. For instance, when adding a rule that marks *all* third person pronouns as referring to documents, precision decreases dramatically, with insufficient increase in recall:  $R = 0.71$ ,  $P = 0.52$  and  $f = 0.60$ . Similarly, adding a rule that marks *all* indexicals as referring to documents produces an even lower performance:  $R = 0.70$ ,  $P = 0.46$  and  $f = 0.56$ . It appears however that in for the present meeting genre, the indexicals ‘celui-ci’ and ‘celui-là’ (‘this one’ and ‘that one’, masculine forms) are almost always used to refer to articles. Therefore, the best scores are obtained after tuning and adding the previous rule:  $R = 0.68$ ,  $P = 0.88$  and  $f = 0.76$ . However, even without this particular rule, f-measure after tuning the grammar is only 1% lower.

## 5.2 Scores for the Resolution of REs

Several baseline scores can be proposed for comparison purposes, depending on the choice of a “minimal” algorithm. For the RE/document association metric, always choosing the most frequent newspaper leads to ca. 80% baseline accuracy. However, when considering meetings with at least two newspapers, the score of this random procedure is 50%, a much more realistic, and lower, baseline. Regarding the RE/element association metric, if the referent is always the front page as a whole, then accuracy is 16%. If the referent is always the main article, then accuracy is 18% – in both cases quite a low baseline.

The RE resolution algorithm applied on the set of *correct* REs reaches 97% accuracy for the identification of documents referred to by REs, i.e., 428 REs out of 437 are correctly resolved. The accuracy is 93% if only meetings with several documents are considered. This is a very high score which proves the relevance of the word co-occurrence and anaphora tracking techniques.

The accuracy for document element identification is 67%, that is, 303 REs out of 437 are correctly resolved at the level of document elements. If we consider only REs for which the correct document was previously identified, the accuracy is 68% (301 REs out of 428). This figure is basically the same since most of the RE/document associations are correctly resolved.

The best scores are obtained when only the right context of the RE is considered for matching, i.e. only the words *after* the RE, and not the ones *before* it. Empirically, the optimal number of words to look for in the right context is about ten. Regarding the other optimal parameters, a match between the RE and the title of an article appears to be more important than one involving the right context of the RE and the title, and much more important than matches with the content of the article: optimal weights are about 15 vs. 10 vs. 1. If anaphor tracking is disabled, the accuracy of document element identification drops to ca. 60%. The observation of systematic error patterns could help us improve the algorithm.

## 5.3 Combination of RE Detection and Resolution

When the two modules are combined in a pipeline, their errors cumulate in a way that is *a priori* unpredictable, but which can be assessed empirically as

follows. The best configurations were selected for the two modules and, on a perfect transcript, the obtained results were: 60% document accuracy (265 REs out of 437) and 32% document element accuracy (141 REs out of 437). If we compute document element accuracy only on the REs which have the correct document attached, the score increases to 46% (123 REs out of 265). It appears thus that the error rates do not combine linearly: if they did, the scores would have been, respectively, ca. 73% and ca. 50%.

The reason for the lower than expected scores lies probably in the context-based algorithm used for RE resolution, in which each RE depends on the correct resolution of the previous one, through the monitoring of the “current document element”. This is a pertinent feature when REs are correctly detected, but when too many REs are missing (here recall is only 67%), and especially when most of the pronouns are missing, the algorithm loses track of the current document element. Therefore, an improvement of the RE detector should considerably increase the overall detection-plus-resolution accuracy.

## 6 Other Document/Speech Alignment Techniques

The resolution of references to documents is not the only method for the cross-channel alignment of meeting dialogues with meeting documents. We have implemented and evaluated two other methods: *citation-based alignment*, a pure lexical match between terms in documents and in speech transcription, and *thematic alignment*, a semantic match between sections of documents (sentences, paragraphs, logical blocks, etc.) and units of dialogue structure (utterances, turns, and thematic episodes).

The robust thematic alignment method uses various state-of-the-art metrics (cosine, Jaccard, Dice), considering document and speech units as bags of weighted words [11]. After suppression of stop-words, proper stemming, and after calculation of terms frequency in their section relative to their frequency in the whole document (TF.IDF), the content of various types of document elements is compared with the content of various speech transcript units.

When matching spoken *utterances* with document *logical blocks*, using cosine metric, recall is 84%, and precision is 77%, which are encouraging results. And when matching speech *turns* with logical blocks, recall stays at 84% and precision rises to 85%. On the other hand, alignment of spoken *utterances* to document *sentences* is less precise but is more promising since it relies on less processing. Using Jaccard metric, recall is 83%, and precision is 76% [11]. Furthermore, thematic alignment of spoken *utterances* to document *sentences* has been used for joint thematic segmentation of documents and speech transcripts. The evaluation of this method shows that this bi-modal thematic segmentation outperforms standard mono-modal segmentation methods, which tends to prove that the combination of modalities considerably improves segmentation scores [12].

In another recent, integrative evaluation, we measured the effect of combining the various document/speech alignments (reference-based, citation-based, and thematic) on the general document/speech alignment performance [13]. Eight

meetings were tested, with a total of 927 utterances, and 116 document logical blocks. After combination of the three methods, the values of recall, precision, and f-measure were respectively 67%, 72% and 68%, whereas their independent use reaches at best, respectively, 55%, 75% and 63%. These results tend to prove the benefit of combining the various methods of document/speech alignment.

## 7 Conclusion

Printed documents and spoken interaction are two important modalities in communication. This article presented an attempt to align these modalities based on their semantic content, in the context of a meeting browser that makes use of the mentions of documents in the dialogue.

The results presented here demonstrate the feasibility of a reference-based alignment technique using a grammar-based module for RE detection, followed by a module implementing word co-occurrence and anaphora tracking for RE resolution. The two modules were evaluated separately, then in sequence: the scores for the overall task remain still above the baseline when the two modules are combined. Future feasibility studies could also evaluate the degradation induced in a pipelined alignment process by other automated modules, such as speech recognition or document structuring. Together with other alignment techniques, we believe that our approach will contribute to the design of a robust multi-modal meeting browser.

## Acknowledgements

This work is part of (IM)2, a project supported by the Swiss National Science Foundation (see <http://www.im2.ch>). The authors would like to thank Dalila Mekhaldi, Emmanuel Palacio and Didier von Rotz for help with data preparation, as well as the three anonymous MLMI'05 reviewers for their valuable suggestions.

## References

1. Lisowska, A., Popescu-Belis, A., Armstrong, S.: User query analysis for the specification and evaluation of a dialogue processing and retrieval system. In: LREC'04, Lisbon (2004) 993–996
2. Hadjar, K., Rigamonti, M., Lalanne, D., Ingold, R.: Xed: a new tool for extracting hidden structures from electronic documents. In: Workshop on Document Image Analysis for Libraries, Palo Alto, CA (2004)
3. Mitkov, R.: Anaphora Resolution. Longman, London, UK (2002)
4. Hirschman, L.: MUC-7 Coreference task 3.0. Technical report, MITRE (1997)
5. van Deemter, K., Kibble, R.: On coreferring: Coreference in MUC and related annotation schemes. *Computational Linguistics* **26**(4) (2000) 629–637
6. Popescu-Belis, A.: Evaluation-driven design of a robust reference resolution system. *Natural Language Engineering* **9**(3) (2003) 281–306

7. Popescu-Belis, A., Lalanne, D.: Reference resolution over a restricted domain: References to documents. In: *ACL'04 Workshop on Reference Resolution and its Applications*, Barcelona (2004) 71–78
8. Popescu-Belis, A., Georgescu, M., Clark, A., Armstrong, S.: Building and using a corpus of shallow dialogue annotated meetings. In: *LREC'04*, Lisbon (2004) 1451–1454
9. Simov, K., Simov, A., Ganev, H., Ivanova, K., Grigorov, I.: The CLaRK system: Xml-based corpora development system for rapid prototyping. In: *LREC'04*, Lisbon (2004) 235–238
10. Grosz, B.J., Joshi, A.K., Weinstein, S.: Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* **21**(2) (1995) 203–225
11. Lalanne, D., Mekhaldi, D., Ingold, R.: Talking about documents: revealing a missing link to multimedia meeting archives. In: *Document Recognition and Retrieval XI*, San Jose, CA (2004)
12. Mekhaldi, D., Lalanne, D., Ingold, R.: Using bi-modal alignment and clustering techniques for documents and speech thematic segmentations. In: *CIKM'04*, Washington D.C. (2004)
13. Mekhaldi, D., Lalanne, D., Ingold, R.: From searching to browsing through multi-modal documents linking. In: *ICDAR'05*, Seoul (2005)